

---

Marc DEBONO (dir.) — Corpus numériques, langues  
et sens. Enjeux épistémologiques et politiques.  
Bruxelles : Peter Lang, 2014.

---

**Electronic version**URL: <http://journals.openedition.org/corpus/2722>

ISSN: 1765-3126

**Publisher**

Bases ; corpus et langage - UMR 6039

**Printed version**

Date of publication: 1 January 2015

ISBN: 1 638-9808

ISSN: 1638-9808

**Electronic reference**

« Marc DEBONO (dir.) — Corpus numériques, langues et sens. Enjeux épistémologiques et politiques.  
Bruxelles : Peter Lang, 2014. », *Corpus* [Online], 14 | 2015, Online since 29 August 2017, connection on  
08 September 2020. URL : <http://journals.openedition.org/corpus/2722>

---

This text was automatically generated on 8 September 2020.

© Tous droits réservés

---

## Marc DEBONO (dir.) — Corpus numériques, langues et sens. Enjeux épistémologiques et politiques. Bruxelles : Peter Lang, 2014.

---

- 1 Corpus numériques, langues et sens est un ouvrage collectif intéressant et problématique publié avec le soutien de la DGLFLF.
- 2 Commençons par rendre compte brièvement de l'aspect problématique pour mieux développer les aspects les plus intéressants d'un livre dont le mérite est d'aborder intelligemment des questions actuellement fondamentales et activement débattues : le statut des observables, des données et du sujet en linguistique, la révolution (ou non-révolution) du numérique dans les Sciences du langage et au sein des digital humanities, la construction du sens dans/par les corpus.
- 3 De manière problématique d'abord, en effet, les auteurs se proposent de parler (pour ne pas dire de critiquer) des corpus numériques, et par-là de la linguistique de corpus, comme les hommes d'église glosent sur le péché de chair : sans beaucoup pratiquer. De corpus et de numérique, le lecteur n'en trouvera pas dans l'ouvrage et se reportera, en la matière, à d'autres écrits des auteurs. Point de donnée ici, point de pratique logicielle<sup>1</sup>, point de main salie par le formatage, l'enrichissement ou l'étiquetage de textes numériques ; point surtout d'étude ou d'analyse effective de corpus de données numériques, natives issues d'internet ou non natives issues de scanner. Ces absences omniprésentes dans les contributions des 6 auteurs et le long des 216 pages du livre<sup>2</sup> pourront être jugées comme dirimantes à la crédibilité de la réflexion proposée mais peuvent se comprendre de deux manières. D'abord, les auteurs, sociolinguistes et rompus par ailleurs aux enquêtes de terrain, ne cachent pas ici leur point de vue épistémologique (cf. le sous-titre de l'ouvrage et l'ensemble des contributions), qui semble autoriser une approche surplombante, théorique et finalement platonique des choses et des corpus. Ensuite, la plupart des auteurs, Marc Debono en tête dès l'introduction, milite explicitement pour des Sciences du langage qui s'arracheraient de l'empirisme, du positivisme des données, des usages consignés en corpus pour

réintroduire plus hautement les représentations et la sensibilité des chercheurs dans l'analyse (dès la p. 11 puis tout au long de l'ouvrage) : finalement, en prenant le parti de ne jamais traiter concrètement de corpus, dans leur matérialité linguistique, les auteurs illustrent parfaitement ce positionnement idéal pas seulement platonique mais platonicien.

- 4 Nonobstant, l'ouvrage apparaît comme roboratif notamment par les prises de position courageuses, et par son « argumentaire contre » revendiqué, selon les vertus affirmées de la « productivité du conflit » scientifique (p. 17). Car l'ouvrage est explicitement militant contre l'idée que les corpus numériques puissent révolutionner nos pratiques scientifiques, contre l'affirmation qu'ils puissent prétendre à une forme d'objectivation (notamment quantitative) du langage, contre l'illusion scientiste que porterait en elle la technologisation de nos approches ; contre, au fond, ce que les auteurs ressentent comme l'épistémologie ou idéologie dominantes actuellement en SDL. « L'argumentation pour », quant à elle, apparaît moins percutante puisqu'à l'absence mentionnée d'étude concrète et positive de corpus s'ajoute pour les auteurs la difficulté d'avancer des propositions alternatives précises méthodologiquement au-delà de grands principes herméneutiques qu'il est effectivement important de rappeler (le chercheur est au centre de sa recherche ; « le sens n'est pas totalement contrôlable/prédictible » p. 41 ; le numérique « reste une construction humaine » (p. 51) ; « le travail interprétatif ne peut... incomber qu'au sujet chercheur » (p. 64) ; « la place de l'interprète est importante dans l'activité de signifiante » (p. 135), etc.).
- 5 Ainsi, le chapitre de Marc Debono, « Représentations et traitement des corpus numériques linguistiques. Quid des représentations du chercheur ? », qui succède à son introduction générale, s'inquiète de l'occultation du chercheur et de ses représentations dans les pratiques actuelles en Sciences du langage ; non prise en compte des représentations du chercheur certes pas nouvelle mais rendue particulièrement problématique avec le numérique et la technicisation/industrialisation des pratiques sur big data. Le propos de Marc Debono renoue ainsi avec la critique de l'objectivisme, du scientisme ou encore de « l'hyper-empirisme » (p. 40) que porterait en lui, selon l'auteur, le tout numérique. A contrario, c'est à un éloge des « préjugés » du chercheur au sens de Gadamer, de « l'intuition » de l'analyste au sens de Taylor ou de la « subjectivité expérientielle » que Marc Debono se livre pour promouvoir une herméneutique des textes et du langage dans laquelle le sujet, avec sa pré-compréhension (p. 34) du monde et des textes, serait au commencement et à la fin.
- 6 On lira cet argumentaire serré et documenté de l'auteur qui s'appuie sur une importante bibliographie, et l'on nuancera un point nodal de l'inquiétude de son propos. Pour Marc Debono, la puissance de la machine semble avoir pour corolaire nécessaire « une interprétation considérablement contrainte » (p. 30) ; d'où l'inquiétude de l'herméneute. Or précisément, pour beaucoup de chercheurs, loin de contraindre l'interprétation humaine, l'herméneutique numérique a pour espoir de la libérer. L'outil (ici l'ordinateur) n'a jamais eu d'autre fonction et d'autre conséquence dans les sciences et ailleurs que de faire levier (ici faire levier sur l'interprétation). Pas plus que la révolution Gutenberg et la presse à imprimer n'ont semble-t-il « contraint » l'herméneutique traditionnelle, limité la puissance de la poésie ou bridé la littérature, la révolution numérique n'a de raison objective de le faire. Cinq cents ans après la révolution Gutenberg, la révolution numérique ou l'herméneutique numérique doivent au contraire, à leur tour, révéler les potentialités des textes et des corpus et donner

matière à la subjectivité créatrice de l'homme, notamment en rendant mieux disponible l'intertextualité/interdiscursivité au sein de l'archive universelle qu'approxime extraordinairement aujourd'hui le web, et ceci sous la seule condition de pratiquer internet, manipuler des fichiers XML et quelques logiciels de traitement, exactement de la même manière que l'on a appris, au plus grand bénéfice de la science des textes, au cours des derniers siècles, à maîtriser l'imprimerie, la lecture d'un folio (plus aisé et moins contraignant qu'un rouleau) ou les notes infrapaginales d'un ouvrage papier. Il nous semble que La mesure et le grain de François Rastier, livre récent que l'on pourrait ajouter à la bibliographie de l'ouvrage, constitue justement ce programme : développer une sémantique interprétative (et non formelle ou objectivante comme le craint Marc Debono) grâce à des parcours de lecture numérique démultipliés, faits d'observables requalifiés, d'hypertextualité, de traitements quantitatifs et qualitatifs ou de croisement de sources multimodales<sup>3</sup>.

- 7 Le chapitre écrit par Elatiana Razafimandimbimanana, « Continuité épistémologique au sein de la révolution numérique », apparaît ambitieux par son volume (37 pages). L'auteure ne semble pas vouloir discuter l'ampleur de la révolution numérique, mais, à défaut, en conteste les conséquences épistémologiques pour les Sciences du langage : « Le propos ne consiste pas à réfuter à tout prix le caractère révolutionnaire du numérique » (p. 48) mais voir dans nos pratiques numériques « une non-révolution épistémologique » (p. 49)<sup>4</sup>. Malgré une prudence d'écriture qui complique l'argumentaire, avec l'usage du conditionnel aussi bien pour exprimer des pensées que l'auteure dénonce que pour exposer des pensées que l'auteure approuve/écrit<sup>5</sup>, ce chapitre rappelle pertinemment des vérités et entend dissiper certaines « illusions de l'objectivité numérique » (p. 68 et ss). Comme chez Marc Debono, il s'agit de replacer le sujet avec sa sensibilité au cœur de la science ou en tout cas des Sciences du langage, jusqu'à nuancer les postures seulement rationnelles considérées péjorativement comme des avatars de la pensée objectiviste (p. 75). Car de fait : « De qualité numérique ou pas, les observables demeurent un effet partiel et dépendant des interprétations propres au chercheur » (p. 76). Resterait dès lors à développer un programme de recherche au-delà des lignes « Perspectives » qui concluent l'article, sonnent comme une ode humaniste au sujet et à une subjectivité réhabilitée, mais demanderaient sans doute plus de place pour devenir effective ou opératoire d'un point de vue pratique ou méthodologique.
- 8 Le chapitre rédigé par Lorenzo Bonoli propose « un regard épistémologique sur les logiciels d'analyse textuelle » et apparaît comme moins sceptique que les deux premiers sur l'apport du numérique. L'auteur se décrit modestement comme non spécialiste desdits logiciels mais en perçoit pourtant pertinemment deux sortes : les logiciels de type CAQDAS (Computer Assisted Data Analysis Software) et les logiciels de type statistique. S'attachant surtout aux logiciels « quantitatifs » (quand bien même ceux-ci disposent toujours, complémentirement à leurs fonctions statistiques, des fonctions qualitatives hypertextuelles de retour au texte), l'auteur semble concéder qu'ils puissent avoir une dimension heuristique en ouvrant des possibles encore inconnus à l'interprétation (note de la p. 84) et, à rebours du chapitre précédent, envisage qu'ils puissent définir un « nouveau cadre épistémologique » (p. 86) ou changer « profondément notre rapport au texte » (p. 91). En conclusion, l'auteur semble également reconnaître « l'apport important et intéressant que les logiciels d'aide à l'analyse textuelle fournissent, tout en restant dans le paradigme interprétatif » (p. 93, nous soulignons). De lui-même, Lorenzo Bonoli cite ainsi Max

Reinert, un des concepteurs pionniers des logiciels en question : « Ce ne sont pas des instruments de validation mais des aides à la construction d'hypothèses, ou même plus simplement des aides à la lecture, car on doit recomposer soi-même sa tresse du sens » (p. 87) ou Roel Popping qui milite pour l'approche logicielle en n'accordant à l'analyse automatique qu'une part raisonnée dans le processus interprétatif (p. 90).

- 9 Aussi peut-être est-ce par projection que l'auteur finit par trouver dans les logiciels qu'il vise, dont *Alceste* de Max Reinert, ce qu'il redoute d'y trouver : « l'évacuation » et même la « double évacuation » (p. 89) du chercheur dans l'interprétation. A moins que le propos ne cible, pertinemment, l'usage caricatural, et nous l'espérons marginal, de certains chercheurs qui confondent encore aujourd'hui, de fait, une sortie-machine et une probation.
- 10 Isabelle Pierozak se propose de traiter « Corpus et numérique en sciences du langage : enjeux épistémologiques ». S'appuyant sur un écrit précédent, elle discute « l'eldorado » (p. 97) que représentent aujourd'hui les données électroniques en SDL, et retrace les différentes définitions données aux corpus d'un point de vue épistémologique, pour s'interroger finalement : « Des corpus numériques pour quoi faire ? » (p. 101). Ses conclusions rejoignent en grande partie celles de Marc Debono et Elatiana Razafimandimbimanana. Pour elle, les corpus numériques sont le plus souvent un « leurre » (p. 102) en ce qu'ils semblent donner à voir des données objectives là où ils ne sont en vérité que des artefacts, trop souvent, selon elle, impensés. Qu'il nous soit permis ici de faire une remarque sur ce chapitre comme sur l'ensemble de l'ouvrage. En rappelant à juste titre la dimension construite des données numériques, pour en réfuter le caractère objectif et par là dénoncer le « néo-néo-positivisme » (selon Lorenzo Bonoli, p. 91) des chercheurs qui s'y aveugleraient, l'ouvrage ne risque-t-il pas de laisser entendre, en creux et évidemment malgré lui, que les données antérieures, non-numériques donc, seraient, elles, naturelles ? En regrettant la « naturalisation » des corpus numériques et les dangers que représente cette naturalisation machinale, n'y a-t-il pas un risque d'essentialiser voire fétichiser les corpus traditionnels et le texte papier d'origine ? Admettons donc que pas plus, mais pas moins, que le texte ou le corpus papier, les corpus numériques sont des objets naturels, et ils ne peuvent prétendre au statut d'absolu. Les maux dont on les accable dans cet ouvrage (empirisme impensé, positivisme des données, scientisme) ne leurs sont ni exclusifs ni définitionnels, et ils n'en sont pas responsables ni historiquement ni épistémologiquement. Du reste, les auteurs en conviennent : Comte, ses épigones et ses critiques précèdent de plusieurs décennies l'invention du numérique, et Bacon de plusieurs siècles<sup>6</sup>.
- 11 Le chapitre de Valentin Feussi traite des « Pratiques linguistiques numériques/électroniques. Une source d'angoisse pour les linguistes ? ». Après avoir établi le *distinguo* terminologique entre « numérique » et « électronique » (admettons effectivement que les deux termes sont le plus souvent quasi synonymes sous la plume des chercheurs), l'auteur pousse plus loin encore l'inquiétude de ses co-auteurs en discutant « l'angoisse du chercheur-linguiste » (p. 123) face/à cause du numérique. Le propos général de l'ouvrage se répète et se précise. Pour Valentin Feussi aussi « l'objectif en tout cas est de privilégier l'interprétation (à l'explication et à la démonstration) » (p. 126). La posture interprétative (versus objectivante) est donc réaffirmée. Mais l'auteur précise l'herméneutique qu'il souhaite et regrette celle que semblent proposer les pratiques numériques. C'est l'herméneutique matérielle

numérique – celle qui s'appuie sur la matérialité des données et prend en considération, selon Rastier, les « contraintes linguistiques sur l'interprétation » (p. 127) – qui est alors assimilée à une « rationalité moderniste » ou à une rançon payée à la « postmodernité » (p. 127). L'auteur regrette ainsi la « technologisation des interprétations » (p. 127) et y oppose une « interprétation humaine » (p. 134) ou « interprétation philosophique » (p. 138) dans laquelle « le primat » est clairement accordé « à l'interpréteur, au chercheur et non à ses outils techniques » (p. 138) ou à son corpus dans sa composition linguistique numérique. En d'autres termes, face aux « utopies technologiques » (p. 138), l'auteur entend développer une utopie du sentiment « en considérant la subjectivité et la sensibilité du chercheur comme éléments structurant dans la production de connaissances scientifiques » (p. 140).

- 12 En guise de conclusion enfin, le lecteur pourra s'attacher aux propos de Didier de Robillard, « Monnaie de signe, monnaie de singe ? Comment comprendre des corpus électroniques ? Implications épistémologiques, éthiques et politiques ». Ce chapitre très conséquent de 70 pages apparaît tangentiel à la problématique avant tout épistémologique de l'ouvrage et développe surtout l'aspect « politique » et institutionnel d'une recherche actuellement dominée par les appels d'offres autour des corpus. C'est notamment « les politiques mises en œuvre au nom de la République française, par la DGLFLF » (p. 146) qui sont alors discutées, avant de conclure puissamment sur les objectifs de l'ouvrage dans ses termes clefs : « Une conception positiviste du corpus » (p. 192), « La marginalisation de la notion de représentation » (p. 194), « Corpus (numérique) et sens » (p. 201).
- 13 Parce qu'il est ouvertement critique sur la situation scientifique actuelle, sans verser pour autant dans la polémique, *Corpus numériques, langues et sens* est un ouvrage utile. Le point de vue engagé des auteurs en faveur de la présence du sujet herméneutique dans les sciences humaines est réinvesti après Dilthey ou Gadamer dans une critique du numérique et de ses corpus qui sans doute n'en méritaient pas tant. Dans une vision inquiète des sciences humaines (forcément menacées par les sciences de la nature) et des techniques (menaçant nécessairement la créativité/liberté humaines), les auteurs minimisent la révolution numérique tout en maximisant ses conséquences négatives. On ne sera donc pas obligé d'adhérer aux idées pessimistes principales de l'ouvrage que le numérique entraîne inévitablement une démarche positiviste non-herméneutique, que l'outil aliène l'homme en contraignant sa liberté ou que, par ailleurs, l'interprétation peut faire l'économie d'une description du matériel linguistique des corpus (Aux limites de l'interprétation, Umberto Eco, lui-même, ne finit-il pas par concéder que pour qu'il y ait interprétation textuelle il faut qu'il y ait d'abord un texte à interpréter ? C'est ce texte-à-interpréter que les corpus (numériques) tentent de formaliser). Mais, il n'en reste pas moins vrai que l'expansion si forte et si rapide du numérique demande à être réfléchie et que l'essor décennal des corpus méritait ce temps d'une réflexion épistémologique que les auteurs nous offrent.
- 14 Damon Mayaffre  
UMR 7320, Bases, Corpus, Langage
- 15 CNRS – Université de Nice Sophia Antipolis

---

## NOTES

1. Lorenzo Bonoli propose cependant une courte typologie des logiciels existants pages 84-85.
2. Valentin Feussi mentionne cependant en haut de la page 139 des données issues d'internet.
3. Poussons la comparaison historique. L'inquiétude sociale et épistémologique actuelle face à la révolution numérique, qui est le sentiment conducteur de l'ensemble de l'ouvrage, semble reprendre l'inquiétude métaphysique passée de l'institution catholique face à la révolution Gutenberg. L'impression démultipliée de la bible (en langue vernaculaire), et sa diffusion de masse dans les foyers chrétiens, semblaient devoir, craignait-on, affaiblir le sentiment et la foi : pas plus que le croyant réformé n'a perdu sa foi en disposant du Livre imprimé – certains se trouvant même raffermis dans leur croyance –, l'herméneute numérique ne devrait perdre – au contraire – sa toute-puissance interprétative en surfant sur le web.
4. Plus radicalement encore : « Plus qu'un non-événement, je dirais que la réalité de la révolution numérique serait quelque part entre le mythe et la rhétorique » (p. 59).
5. Par exemple « le numérique serait à même de rendre la complexité et l'opacité des phénomènes » : ici on comprend que l'auteure réfute ce point de vue jugé comme naïf. Mais après « la révolution numérique n'aurait pas [...] eu de retentissements significatifs en termes de connaissances/compréhensions des fonctionnements des langues » (p. 62) : ici c'est toute la position de l'auteure qui semble mise sous condition(el). L'ensemble des pages 62 et 63 notamment sont ainsi au conditionnel pour exprimer ce que l'on a cru comprendre être les convictions fortes de l'auteure.
6. C'est Lorenzo Bonoli qui décrit ainsi la pièce que l'ouvrage semble mettre en scène : « La tentative de leur introduction [ie. *grosso modo* l'introduction des instruments et pratiques numériques] dans le domaine des disciplines du texte, et plus généralement des SHS, semble relancer un nouveau positivisme, une sorte de néo-néo-positivisme qui oblige les chercheurs dans les SHS [ie. en tout cas les auteurs de l'ouvrage] à défendre des spécificités de leur domaine de recherche en reprenant finalement les arguments que leurs prédécesseurs avaient déjà dû évoquer contre le positivisme du XIX<sup>e</sup> et le néo-positivisme du début du XX<sup>e</sup> siècle » (p. 92).